

SEQUENCE LISTING

5 (1) GENERAL INFORMATION

(i) APPLICANT

NAME: F. HOFFMANN-LA ROCHE AG

STREET: Grenzacherstrasse 124

10 CITY: Basle

COUNTRY: Switzerland

POSTAL CODE: CH-4002

TELEPHONE: 061 - 688 25 05

FAX: 061 - 688 13 95

15 TELEX: 962292/965542 hlr c

(ii) TITLE OF INVENTION:

Novel Alcohol/Aldehyde Dehydrogenases

(iii) NUMBER OF SEQUENCES: 12

(iv) COMPUTER READABLE FORM:

20 (A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: Macintosh

(C) OPERATING SYSTEM:

(D) SOFTWARE: MS word ver 5.1

25

{ 65 } replaced with
4/21/98 Amendment

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1740 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) ORIGINAL SOURCE:

ORGANISM: *Gluconobacter oxydans*

10 STRAIN: DSM 4025

(iv) FEATURE:

FEATURE KEY: CDS

POSITION: 1..1737

SEQUENCING METHOD: E

15 ATGAAACCGA CTTCGCTGCT TTGGGCCAGT GCTGGCGCAC TTGCATTGCT 50
20 TGCCGCACCC GCCTTGCTC AAGTGACCCC CGTCACCGAT GAATTGCTGG 100
CGAACCCGCC CGCTGGTGAA TGGATCAGCT ACGGTCAGAA CCAAGAAAAC 150
25 TACCGTCACT CGCCCCTGAC GCAGATCAGG ACTGAGAACG TCGGCCAACT 200
GCAACTGGTC TGGGCGCGCG GCATGCAGCC GGGCAAAGTC CAAGTCACGC 250
CCCTGATCCA TGACGGCGTC ATGTATCTGG CAAACCCGGG CGACGTGATC 300
30 CAGGCCATCG ACGCCAAAAC TGGCGATCTG ATCTGGAAC ACCGCCGCCA 350
ACTGCCGAAC ATCGCCACGC TGAACAGCTT TGGCGAGCCG ACCCGCGGCA 400
TGGCGCTGTA CGGCACCAAC GTTTACTTTG TTTCGTGGGA CAACCACCTG 450
35 GTCGCCCTCG ACACCGCAAC TGGCCAAGTG ACGTTCGACG TCGACCGCGG 500
CCAAGGCGAA GACATGGTTT CGAACTCGTC GGGCCCGATC GTGGCAAACG 550

5 GCGTGATCGT TGCCGGTTCG ACCTGCCAAT ACTCGCCGTT CGGCTGCTT 600
GTCTCGGGCC ACGACTCGGC CACCGGTGAA GAGCTGTGGC GCAACTACTT 650
CATCCCGCCG GCTGGCGAAG AGGGTGATGA GACTTGGGGC AACGATTACG 700
AAGCCCGTTG GATGACCGGT GCCTGGGCC AGATCACCTA TGACCCCGTC 750
10 ACCAACCTTG TCCACTACGG CTCGACCGCT GTGGGTCCGG CGTCGGAAAC 800
CCAACGCGGC ACCCCGGGCG GCACGCTGTA CGGCACGAAC ACCCGTTTCG 850
15 CGGTGCGTCC TGACACGGGC GAGATTGTCT GGCGTCACCA GACCCTGCC 900
CGCGACAACT GGGACCAGGA ATGCACGTTG GAGATGATGG TCACCAATGT 950
GGATGTCCAA CCCTCGACCG AGATGGAAGG TCTGCAGTCG ATCAACCCGA 1000
20 ACGCCGCAAC TGGCGAGCGT CGCGTGCTGA CCGCGTTCC GTGCAAAACC 1050
GGCACCATGT GGCAGTTCGA CGCCGAAACC GGCGAATTCC TGTGGGCCCG 1100
25 TGATACCAAC TACCAGAACATGATCGAACATC CATCGACGAA AACGGCATCG 1150
TGACCGTGAA CGAAGATGCG ATCCTGAAGG AACTGGATGT TGAATATGAC 1200
GTCTGCCCGA CCTTCTTGGG CGGCCGCGAC TGGCCGTCGG CCGCACTGAA 1250
30 CCCCGACAGC GGCATCTACT TCATCCCGCT GAACAACGTC TGCTATGACA 1300
TGATGGCCGT CGATCAGGAA TTCACCTCGA TGGACGTCTA TAACACCAGC 1350
AACGTGACCA AGCTGCCGCC CGGCAAGGAT ATGATCGGTC GTATTGACGC 1400
35 GATCGACATC AGCACGGGTC GTACGCTGTG GTCGGTCGAA CGTGCTGCGG 1450
CGAACTATTG GCCCGTCTTG TCGACCGGCG GCGGCGTTCT GTTCAACGGT 1500
40 GGTACGGATC GTTACTTCCG CGCCCTCAGC CAAGAAACCG GCGAGACCCCT 1550
GTGGCAGACC CGCCTTGCAA CCGTCGCGTC GGGCCAGGCC ATCTCTTACG 1600
45 AGGTTGACGG CATGCAATAT GTCGCCATCG CAGGTGGTGG TGTCAGCTAT 1650
GGCTCGGGCC TGAACCTCGGC ACTGGCTGGC GACCGAGTCG ACTCGACCGC 1700
CATCGGTAAC GCCGTCTACG TCTTCGCCCT GCCGCAATAA 1740

INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1740 base pairs
(B) TYPE: nucleic acid
5 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) ORIGINAL SOURCE:

ORGANISM: *Gluconobacter oxydans*

10 STRAIN: DSM 4025

(iv) FEATURE:

FEATURED KEY: CDS

POSITION: 1..1737

15 SEQUENCING METHOD: E

ATGAAGACGT CGTCTTGCT GGTTCCGAGC GTTGCAGCGC TTGCAAGCTA 50
20 TAGCTCCTTT GCGCTTGCTC AAGTGACCCCC CGTCACCGAT GAATTGCTGG 100
CGAACCCGCC CGCTGGTGAA TGGATCAGCT ACGGTCAGAA CCAAGAAAAC 150
25 TACCGTCACT CGCCCCTGAC GCAGATCAG ACTGAGAACG TCGGCCAACT 200
GCAACTGGTC TGGGCGCGCG GCATGCAGCC GGGCAAAGTC CAAGTCACGC 250
CCCTGATCCA TGACGGCGTC ATGTATCTGG CAAACCCGGG CGACGTGATC 300
30 CAGGCCATCG ACGCCAAAAC TGGCGATCTG ATGTGGGAAC ACCGCCGCCA 350
ACTGCCGAAC ATCGCCACGC TGAACAGCTT TGGCGAGCCG ACCCGCGGCA 400
35 TGGCGCTGTA CGGCACCAAC GTTTACTTTG TTTCTGGGGAA CAACCACCTG 450
GTCGCCCTCG ACACCGCAAC TGGCCAAGTG ACGTTCGACG TCGACCGCGG 500
CCAAGGCGAA GACATGGTTT CGAACTCGTC GGGCCCGATC GTGGCAAACG 550

50
5 GCGTGATCGT TGCCGGTTCG ACCTGCCAAT ACTGCCGTT CGGCTGCTT 600
GTCTCGGGCC ACGACTCGGC CACCGGTGAA GAGCTGTGGC GCAACTACTT 650
CATCCCGCGC GCTGGCGAAG AGGGTGATGA GACTTGGGGC AACGATTACG 700
AAGCCCGTTG GATGACCGGC GTCTGGGTC AGATCACCTA TGACCCCGTT 750
10 GGCGGCCTTG TCCACTACGG CTCGTGGCT GTTGGCCCGG CTTCGGAAAC 800
CCAGCGCGC ACCACCGGCG GCACCATGTA CGGCACCAAC ACCCGTTTCG 850
15 CTGTCCGTCC CGAGACTGGC GAGATCGTCT GGCGTCACCA AACTCTGCC 900
CGCGACAACG GGGACCAAGA GTGCACCTTC GAGATGATGG TTGCCAACGT 950
TGACGTGCAG CCCGCAGCTG ACATGGACGG CGTCCGCTCG ATCAACCCGA 1000
20 ACGCCGCCAC CGGCGAGCGT CGCGTTCTGA CCGGCGTTCC GTGAAAACC 1050
GGCACCATGT GGCAGTTCGA CGCCGAAACC GGCGAATTCC TGTGGCCCG 1100
25 TGACACCAGC TACGAGAACAC TCATCGAATC GATCGACGAA AACGGCATCG 1150
TGACCGTCGA CGAGTCGAAA GTTCTGACCG AGCTGGACAC CCCCTATGAC 1200
GTCTGCCCGC TGCTGCTGGG TGGCCGTGAC TGGCCGTCGG CTGCGCTGAA 1250
30 CCCCCGATACC GGCATCTACT TTATCCCGCT GAACAACACC TGCATGGATA 1300
TCGAAGCTGT CGACCAGGAA TTCAGCTCGC TGGACGTGTA CAACCAAAGC 1350
CTGACCGCCA AAATGGCACC GGGTAAAGAG CTGGTTGGCC GTATGACGC 1400
35 CATCGACATC AGCACAGGCC GCACCCCTGTG GACCGCTGAG CGCGAAGCCT 1450
CGAACTACGC GCCTGTCCTG TCGACCGCTG GCGGCGTTCT GTTCAACGGC 1500
40 GGCACCGACC GTTACTTCCG CGCTCTCAGC CAAGAGACCG GCGAGACCCCT 1550
GTGGCAGACC CGTCTGGCGA CTGTCGCTTC GGGCCAAGCT GTCTCGTACG 1600
45 AGATCGACGG CGTCCAATAC ATCGCCATCG GCGGGGGGGG CACGACCTAT 1650
GGTCGTTCC ACAACCGTCC CCTGGCCGAG CCGGTGACT CGACCGCGAT 1700
CGGTAATGCG ATGTACGTCT TCGCGCTGCC CCAGCAATAA 1740

50

INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1737 base pairs
5 (B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) ORIGINAL SOURCE:

10 ORGANISM: *Gluconobacter oxydans*

STRAIN: DSM 4025

(iv) FEATURE:

FEATURE KEY: CDS

POSITION: 1..1734

15 SEQUENCING METHOD: E

20 ATGAAACTGA CGACCCTGCT GCAAAGCAGC GCCGCCCTGC TTGTGCTTGG 50
CACCATTCCC GCCCTTGCCC AAACCGCCAT CACCGATGAA ATGCTGGCGA 100
ACCCGCCCGC TGGTGAATGG ATCAACTACG GTCAGAACCA AGAGAACTAC 150
25 CGCCACTCGC CCCTGACGCA GATTACCGCA GACAACGTCG GCCAACTGCA 200
ACTGGTCTGG GCGCGCGGTA TGGAAGCGGG CAAGATCAA GTGACCCCCG 250
30 TTGTCCATGA CGGCGTCATG TATCTGGCAA ACCCCGGTGA CGTGATCCAG 300
GCCATCGACG CCGCGACCGG CGATCTGATC TGGGAACACCC GCCGCCAACT 350
GCCGAACATC GCCACGCTGA ACAGCTTGG TGAGCCGACC CGCGGCATGG 400
35 CCCTCTATGG CACCAACGTC TATTTCGTCT CGTGGGACAA CCACTTGGTC 450
GCGCTGGACA CCTCGACCGG CCAAGTCGTA TTTCGACGTCG ATCGCGGTCA 500

AGGCACGGAT ATGGTCTCGA ACTCGTCCGG CCCGATTGTC GCCAATGGCG 550
5 TCATCGTTGC GGGCTCGACC TGTCAGTATT CGCCGTTCGG CTGTTTCGTT 600
TCGGGCCACG ACTCGGCCAC CGGTGAAGAG CTGTGGCGCA ACAACTTTAT 650
CCCAGCGGCC GGCGAAGAGG GTGATGAGAC CTGGGGCAAT GATTACGAGG 700
10 CCCGCTGGAT GACCGCGT TGGGGCCAGA TCACCTATGA CCCCGTTGGC 750
GGCCTTGTCC ACTACGGCAC CTCAGCAGTT GGCCCTGCGG CCGAGATTCA 800
15 GCGCGGCACC GTTGGCGGCT CGATGTATGG CACCAACACC CGCTTGCTG 850
TCCGCCCGA GACCGCGAG ATCGTCTGGC GTCACCAAAC TCTGCCCGC 900
GACAACACTGGG ACCAAGAGTG TACGTTCGAG ATGATGGTCG TCAACGTCGA 950
20 CGTCCAGCCC TCGGCTGAGA TGGAAAGGCCT GCACGCCATC AACCCCGATG 1000
CCGCCACGGG CGAGCGTCGC GTTGTGACCG GCGTTCCGTG CAAGAACGGC 1050
25 ACCATGTGGC AGTCGACGC CGAAACCGGC GAATTCCCTGT GGGCGCGCGA 1100
CACCAAGCTAT CAGAACCTGA TCGAAAGCGT CGATCCCGAT GGTCTGGTGC 1150
ATGTGAACGA AGATCTGGTC GTGACCGAGC TGGAAAGTGGC CTATGAAATC 1200
30 TGCCCGACCT TCCTGGGTGG CCGCGACTGG CCGTCGGCTG CGCTGAACCC 1250
CGATACTGGC ATCTATTCA TCCCGCTGAA CAACGCCTGT AGCGGTATGA 1300
35 CGGCTGTCGA CCAAGAGTTC AGCTCGCTCG ATGTGTATAA CGTCAGCCTC 1350
GACTATAAAC TGTCGCCCCGG TTCCGAAAAC ATGGGCCGTA TCGACGCCAT 1400
CGACATCAGC ACCGGCCGCA CGCTGTGGTC GGCTGAACGC TACGCCCTGA 1450
40 ACTACGCGCC TGTCCCTGTCC ACCGGCGCG GCGTGCTGTT CAACGGCGGC 1500
ACCGACCGTT ACTTCCGCGC CCTCAGCGAA GAGACCGCG AGACGCTGTG 1550
GCAGACCCGT CTGGCGACTG TCGCCTCGGG TCAAGCGATT TCCTATGAGA 1600
45 TCGACGGCGT GCAATATGTC GCCATCGGGC GCGGCGGCAC CAGCTATGGC 1650
AGCAACCACA ACCGCGCCCT GACCGAGCGG ATCGACTCGA CCGCCATCGG 1700
50 CAGCGCGATC TATGTCTTTG CTCTGCCGCA GAGTAA 1737

INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1740 base pairs
5 (B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) ORIGINAL SOURCE:

10 ORGANISM: *Gluconobacter oxydans*

STRAIN: DSM 4025

(iv) FEATURE:

FEATURE KEY: CDS

POSITION: 1..1737

15 SEQUENCING METHOD: E

20 ATGAACCCCA CAACGCTGCT TCGCACCAAGC GCGGCCGTGC TATTGCTTAC 50
CGCGCCCGCC GCATTCGCGC AGGTAACCCC GATTACCGAT GAACTGCTGG 100
CGAACCCGCC CGCTGGTGAA TGGATTAACT ACGGCCGCAA CCAAGAAAAC 150
TATCGCCACT CGCCCCCTGAC CCAGATCACT GCCGACAACG TTGGTCAGTT 200
GCAACTGGTC TGGGCCCGCG GGATGGAGGC GGGGGCCGTA CAGGTCACGC 250
30 CGATGATCCA TGATGGCGTG ATGTATCTGG CAAACCCCGG TGATGTGATC 300
CAGGCGCTGG ATGCGCAAAC AGGCGATCTG ATCTGGGAAC ACCGCCGCCA 350
ACTGCCCGCC GTGCCACGC TAAACGCCCA AGGCGACCGC AAGCGCGGCG 400
35 TCGCCCTTTA CGGCACGAGC CTCTATTCA GCTCATGGGA CAACCATCTG 450
ATCGCGCTGG ATATGGAGAC GGGCCAGGTC GTATTGATG TCGAACGTGG 500

ATCGGGCGAA GACGGCTTGA CCAGTAACAC CACGGGGCCG ATTGTCGCCA 550
ATGGCGTCAT CGTCGCGGGT TCCACCTGCC AATATTGCC CTATGGATGC 600
5 TTTATCTCGG GGCACGATTG CGCGACGGGT GAGGAGCTGT GCGCAACCA 650
CTTTATCCCG CAGCCGGCG AAGAGGGTGA CGAGACTTGG GGCAATGATT 700
10 TCGAGGCGCG CTGGATGACC GGCCTCTGGG GTCAGATCAC CTATGATCCC 750
GTGACGAACC TTGTGTTCTA TGGCTCGACC GGCCTGGGCC CAGCGTCCGA 800
15 AACCCAGCGC GGCACGCCGG GCGGCACGCT GTATGGCACC AACACCCGCT 850
TTGCGGTGCG TCCCGACACG GGCGAGATTG TCTGGCGTCA CCAGACCCCTG 900
CCGCGCGACA ACTGGGACCA AGAATGCACG TTCGAGATGA TGGTCGCCAA 950
20 CGTCGATGTG CAACCCTCGG CGAGATGGA GGGTCTGCGC GCCATCAACC 1000
CCAATGCGGC GACGGGCGAG CGCCGTGTGC TGACGGGTGC GCCTGCAAG 1050
25 ACCGGCACGA TGTGGTCGTT TGATGCGGCC TCGGGCGAAT TCCTGTGGC 1100
GCGTGATACC AACTACACCA ATATGATCGC CTCGATCGAC GAGACCGGCC 1150
TTGTGACGGT GAACGAGGAT GCGGTGCTGA AAGAGCTGGA CGTTGAATAT 1200
30 GACGTCTGCC CGACCTTCCT GGGTGGGCC GACTGGTCGT CAGCCGCACT 1250
GAACCCGGAC ACCGGCATT ACTTCTGCC GCTGAACAAT GCCTGCTACG 1300
35 ATATTATGGC CGTTGATCAA GAGTTAGCG CGCTCGACGT CTATAACACC 1350
AGCGCGACCG CAAAACTCGC GCCGGGCTTT GAAAATATGG GCCGCATCGA 1400
CGCGATTGAT ATCAGCACCG GGCGCACCTT GTGGTCGGCG GAGCGCCCTG 1450
40 CGGCGAACTA CTCGCCCGTT TTGTCGACGG CAGGCGGTGT GGTGTTCAAC 1500
GGCGGGACCG ACCGCTATTT CCGTGCCCTC AGCCAGGAAA CCGGCGAGAC 1550
TTTGTGGCAG GCGCGTCTTG CGACGGTCGC GACGGGGCAG GCGATCAGCT 1600
45 ACGAGTTGGA CGGCGTGCAA TATATGCCA TCGGTGCGGG CGGTCTGACC 1650
TATGGCACGC AATTGAACGC GCCGCTGGCC GAGGCAATCG ATTGACCTC 1700
50 GGTCGGTAAT GCGATCTATG TCTTGCACT GCCGCAGTAA 1740

INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 579 residues

(B) TYPE: amino acid

5 (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) ORIGINAL SOURCE:

ORGANISM: *Gluconobacter oxydans*

10 STRAIN: DSM 4025

15 (iv) FEATURE:

FEATURE KEY: sig peptide

POSITION: -23..-1

15 SEQUENCING METHOD: E

20 FEATURE KEY: mat peptide

POSITION: 1..556

25 SEQUENCING METHOD: E

Met Lys Pro Thr Ser Leu Leu Trp Ala Ser Ala Gly Ala Leu Ala
-20 -15 -10

Leu Leu Ala Ala Pro Ala Phe Ala Gln Val Thr Pro Val Thr Asp
-5 1 5

Glu Leu Leu Ala Asn Pro Pro Ala Gly Glu Trp Ile Ser Tyr Gly
10 15 20

35 Gln Asn Gla Glu Asn Tyr Arg His Ser Pro Leu Thr Gln Ile Thr
25 30 35

40 Thr Glu Asn Val Gly Gln Leu Gln Leu Val Trp Ala Arg Gly Met
40 45 50

40 Gln Pro Gly Lys Val Gln Val Thr Pro Leu Ile His Asp Gly Val
55 60 65

Met Tyr Leu Ala Asn Pro Gly Asp Val Ile Gln Ala Ile Asp Ala
70 75 80

5 Lys Thr Gly Asp Leu Ile Trp Glu His Arg Arg Gln Leu Pro Asn
85 90 95

Ile Ala Thr Leu Asn Ser Phe Gly Glu Pro Thr Arg Gly Met Ala
100 105 110

10 Leu Tyr Gly Thr Asn Val Tyr Phe Val Ser Trp Asp Asn His Leu
115 120 125

Val Ala Leu Asp Thr Ala Thr Gly Gln Val Thr Phe Asp Val Asp
15 130 135 140

Arg Gly Gln Gly Glu Asp Met Val Ser Asn Ser Ser Gly Pro Ile
145 150 155

20 Val Ala Asn Gly Val Ile Val Ala Gly Ser Thr Cys Gln Tyr Ser
160 165 170

Pro Phe Gly Cys Phe Val Ser Gly His Asp Ser Ala Thr Gly Glu
25 175 180 185

Glu Leu Trp Arg Asn Tyr Phe Ile Pro Arg Ala Gly Glu Glu Gly
190 195 200

Asp Glu Thr Trp Gly Asn Asp Tyr Glu Ala Arg Trp Met Thr Gly
30 205 210 215

Ala Trp Gly Gln Ile Thr Tyr Asp Pro Val Thr Asn Leu Val His
220 225 230

35 Tyr Gly Ser Thr Ala Val Gly Pro Ala Ser Glu Thr Gln Arg Gly
235 240 245

Thr Pro Gly Gly Thr Leu Tyr Gly Thr Asn Thr Arg Phe Ala Val
40 250 255 260

Arg Pro Asp Thr Gly Glu Ile Val Trp Arg His Gln Thr Leu Pro
265 270 275

45 Arg Asp Asn Trp Asp Gln Glu Cys Thr Phe Glu Met Met Val Thr
280 285 290

Asn Val Asp Val Gln Pro Ser Thr Glu Met Glu Gly Leu Gln Ser
295 300 305

50 Ile Asn Pro Asn Ala Ala Thr Gly Glu Arg Arg Val Leu Thr Gly
310 315 320

Val Pro Cys Lys Thr Gly Thr Met Trp Gln Phe Asp Ala Glu Thr
325 330 335

5 Gly Glu Phe Leu Trp Ala Arg Asp Thr Asn Tyr Gln Asn Met Ile
340 345 350

10 Glu Ser Ile Asp Glu Asn Gly Ile Val Thr Val Asn Glu Asp Ala
355 360 365

Ile Leu Lys Glu Leu Asp Val Glu Tyr Asp Val Cys Pro Thr Phe
370 375 380

15 Leu Gly Gly Arg Asp Trp Pro Ser Ala Ala Leu Asn Pro Asp Ser
385 390 395

Gly Ile Tyr Phe Ile Pro Leu Asn Asn Val Cys Tyr Asp Met Met
400 405 410

20 Ala Val Asp Gln Glu Phe Thr Ser Met Asp Val Tyr Asn Thr Ser
415 420 425

Asn Val Thr Lys Leu Pro Pro Gly Lys Asp Met Ile Gly Arg Ile
430 435 440

Asp Ala Ile Asp Ile Ser Thr Gly Arg Thr Leu Trp Ser Val Glu
445 450 455

30 Arg Ala Ala Ala Asn Tyr Ser Pro Val Leu Ser Thr Gly Gly Gly
460 465 470

Val Leu Phe Asn Gly Gly Thr Asp Arg Tyr Phe Arg Ala Leu Ser
475 480 485

35 Gln Glu Thr Gly Glu Thr Leu Trp Gln Thr Arg Leu Ala Thr Val
490 495 500

Ala Ser Gly Gln Ala Ile Ser Tyr Glu Val Asp Gly Met Gln Tyr
505 510 515

Val Ala Ile Ala Gly Gly Val Ser Tyr Gly Ser Gly Leu Asn
520 525 530

45 Ser Ala Leu Ala Gly Glu Arg Val Asp Ser Thr Ala Ile Gly Asn
535 540 545

Ala Val Tyr Val Phe Ala Leu Pro Gln
550 555

INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 579 residues

(B) TYPE: amino acid

5 (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) ORIGINAL SOURCE:

ORGANISM: *Gluconobacter oxydans*

STRAIN: DSM 4025

10 (iv) FEATURE:

FEATURE KEY: sig peptide

POSITION: -23..-1

15 SEQUENCING METHOD: S

FEATURE KEY: mat peptide

20 POSITION: 1..556

25 SEQUENCING METHOD: S

Met Lys Thr Ser Ser Leu Leu Val Ala Ser Val Ala Ala Leu Ala
-20 -15 -10

30 Ser Tyr Ser Ser Phe Ala Leu Ala Gln Val Thr Pro Val Thr Asp
-5 1 5

Glu Leu Leu Ala Asn Pro Pro Ala Gly Glu Trp Ile Ser Tyr Gly
10 15 20

35 Gln Asn Gln Glu Asn Tyr Arg His Ser Pro Leu Thr Gln Ile Thr
25 30 35

40 Thr Glu Asn Val Gly Gln Leu Gln Leu Val Trp Ala Arg Gly Met
40 45 50

45 Gln Pro Gly Lys Val Gln Val Thr Pro Leu Ile His Asp Gly Val
55 60 65

Met Tyr Leu Ala Asn Pro Gly Asp Val Ile Gln Ala Ile Asp Ala
70 75 80

5 Lys Thr Gly Asp Leu Ile Trp Glu His Arg Arg Gln Leu Pro Asn
85 90 95

Ile Ala Thr Leu Asn Ser Phe Gly Glu Pro Thr Arg Gly Met Ala
100 105 110

10 Leu Tyr Gly Thr Asn Val Tyr Phe Val Ser Trp Asp Asn His Leu
115 120 125

15 Val Ala Leu Asp Thr Ala Thr Gly Gln Val Thr Phe Asp Val Asp
130 135 140

Arg Gly Gln Gly Glu Asp Met Val Ser Asn Ser Ser Gly Pro Ile
145 150 155

20 Val Ala Asn Gly Val Ile Val Ala Gly Ser Thr Cys Gln Tyr Ser
160 165 170

Pro Phe Gly Cys Phe Val Ser Gly His Asp Ser Ala Thr Gly Glu
175 180 185

25 Glu Leu Trp Arg Asn Tyr Phe Ile Pro Arg Ala Gly Glu Glu Gly
190 195 200

30 Asp Glu Thr Trp Gly Asn Asp Tyr Glu Ala Arg Trp Met Thr Gly
205 210 215

Val Trp Gly Gln Ile Thr Tyr Asp Pro Val Gly Gly Leu Val His
220 225 230

35 Tyr Gly Ser Ser Ala Val Gly Pro Ala Ser Glu Thr Gln Arg Gly
235 240 245

40 Thr Thr Gly Gly Thr Met Tyr Gly Thr Asn Thr Arg Phe Ala Val
250 255 260

Arg Pro Glu Thr Gly Glu Ile Val Trp Arg His Gln Thr Leu Pro
265 270 275

45 Arg Asp Asn Trp Asp Gln Glu Cys Thr Phe Glu Met Met Val Ala
280 285 290

Asn Val Asp Val Gln Pro Ala Ala Asp Met Asp Gly Val Arg Ser
295 300 305

50 Ile Asn Pro Asn Ala Ala Thr Gly Glu Arg Arg Val Leu Thr Gly
310 315 320

Sub P1
1970

Val Pro Cys Lys Thr Gly Thr Met Trp Gln Phe Asp Ala Glu Thr
325 330 335

5 Gly Glu Phe Leu Trp Ala Arg Asp Thr Ser Tyr Glu Asn Ile Ile
340 345 350

Glu Ser Ile Asp Glu Asn Gly Ile Val Thr Val Asp Glu Ser Lys
10 355 360 365

10 Val Leu Thr Glu Leu Asp Thr Pro Tyr Asp Val Cys Pro Leu Leu
370 375 380

15 Leu Gly Gly Arg Asp Trp Pro Ser Ala Ala Leu Asn Pro Asp Thr
385 390 395

Gly Ile Tyr Phe Ile Pro Leu Asn Asn Thr Cys Met Asp Ile Glu
20 400 405 410

20 Ala Val Asp Gln Glu Phe Ser Ser Leu Asp Val Tyr Asn Gln Ser
415 420 425

25 Leu Thr Ala Lys Met Ala Pro Gly Lys Glu Leu Val Gly Arg Ile
430 435 440

Asp Ala Ile Asp Ile Ser Thr Gly Arg Thr Leu Trp Thr Ala Glu
25 445 450 455

30 Arg Glu Ala Ser Asn Tyr Ala Pro Val Leu Ser Thr Ala Gly Gly
460 465 470

30 Val Leu Phe Asn Gly Gly Thr Asp Arg Tyr Phe Arg Ala Leu Ser
475 480 485

35 Gln Glu Thr Gly Glu Thr Leu Trp Gln Thr Arg Leu Ala Thr Val
490 495 500

40 Ala Ser Gly Gln Ala Val Ser Tyr Glu Ile Asp Gly Val Gln Tyr
505 510 515

Ile Ala Ile Gly Gly Gly Thr Thr Tyr Gly Ser Phe His Asn
45 520 525 530

45 Arg Pro Leu Ala Glu Pro Val Asp Ser Thr Ala Ile Gly Asn Ala
535 540 545

Met Tyr Val Phe Ala Leu Pro Gln Gln
550 555

INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 578 residues

(B) TYPE: amino acid

5 (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) ORIGINAL SOURCE:

ORGANISM: *Gluconobacter oxydans*

10 STRAIN: DSM 4025

15 (iv) FEATURE:

FEATURE KEY: sig peptide

POSITION: -23..-1

15 SEQUENCING METHOD: S

20 FEATURE KEY: mat peptide

POSITION: 1..555

25 SEQUENCING METHOD: S

Met Lys Leu Thr Thr Leu Leu Gln Ser Ser Ala Ala Leu Leu Val
-20 -15 -10

Leu Gly Thr Ile Pro Ala Leu Ala Gln Thr Ala Ile Thr Asp Glu
-5 1 5

Met Leu Ala Asn Pro Pro Ala Gly Glu Trp Ile Asn Tyr Gly Gln
10 15 20

35 Asn Gln Glu Asn Tyr Arg His Ser Pro Leu Thr Gln Ile Thr Ala
25 30 35

Asp Asn Val Gly Gln Leu Gln Leu Val Trp Ala Arg Gly Met Glu
40 45 50

40 Ala Gly Lys Ile Gln Val Thr Pro Leu Val His Asp Gly Val Met
55 60 65

part 1
part 2

	Tyr	Leu	Ala	Asn	Pro	Gly	Asp	Val	Ile	Gln	Ala	Ile	Asp	Ala	Ala
		70						75					80		
5	Thr	Gly	Asp	Leu	Ile	Trp	Glu	His	Arg	Arg	Gln	Leu	Pro	Asn	Ile
			85					90					95		
10	Ala	Thr	Leu	Asn	Ser	Phe	Gly	Glu	Pro	Thr	Arg	Gly	Met	Ala	Leu
			100					105					110		
15	Tyr	Gly	Thr	Asn	Val	Tyr	Phe	Val	Ser	Trp	Asp	Asn	His	Leu	Val
			115					120					125		
20	Ala	Leu	Asp	Thr	Ser	Thr	Gly	Gln	Val	Val	Phe	Asp	Val	Asp	Arg
			130					135					140		
25	Gly	Gln	Gly	Thr	Asp	Met	Val	Ser	Asn	Ser	Ser	Gly	Pro	Ile	Val
			145					150					155		
30	Ala	Asn	Gly	Val	Ile	Val	Ala	Gly	Ser	Thr	Cys	Gln	Tyr	Ser	Pro
			160					165					170		
35	Phe	Gly	Cys	Phe	Val	Ser	Gly	His	Asp	Ser	Ala	Thr	Gly	Glu	Glu
			175					180					185		
40	Leu	Trp	Arg	Asn	Asn	Phe	Ile	Pro	Arg	Ala	Gly	Glu	Glu	Gly	Asp
			190					195					200		
45	Glu	Thr	Trp	Gly	Asn	Asp	Tyr	Glu	Ala	Arg	Trp	Met	Thr	Gly	Val
			205					210					215		
50	Trp	Gly	Gln	Ile	Thr	Tyr	Asp	Pro	Val	Gly	Gly	Leu	Val	His	Tyr
			220					225					230		
55	Gly	Thr	Ser	Ala	Val	Gly	Pro	Ala	Ala	Glu	Ile	Gln	Arg	Gly	Thr
			235					240					245		
60	Val	Gly	Gly	Ser	Met	Tyr	Gly	Thr	Asn	Thr	Arg	Phe	Ala	Val	Arg
			250					255					260		
65	Pro	Glu	Thr	Gly	Glu	Ile	Val	Trp	Arg	His	Gln	Thr	Leu	Pro	Arg
			265					270					275		
70	Asp	Asn	Trp	Asp	Gln	Glu	Cys	Thr	Phe	Glu	Met	Met	Val	Val	Asn
			280					285					290		
75	Val	Asp	Val	Gln	Pro	Ser	Ala	Glu	Met	Glu	Gly	Leu	His	Ala	Ile
			295					300					305		
80	Asn	Pro	Asp	Ala	Ala	Thr	Gly	Glu	Arg	Arg	Val	Val	Thr	Gly	Val
			310					315					320		

Pro Cys Lys Asn Gly Thr Met Trp Gln Phe Asp Ala Glu Thr Gly
325 330 335

5 Glu Phe Leu Trp Ala Arg Asp Thr Ser Tyr Gln Asn Leu Ile Glu
340 345 350

10 Ser Val Asp Pro Asp Gly Leu Val His Val Asn Glu Asp Leu Val
355 360 365

Val Thr Glu Leu Glu Val Ala Tyr Glu Ile Cys Pro Thr Phe Leu
370 375 380

15 Gly Gly Arg Asp Trp Pro Ser Ala Ala Leu Asn Pro Asp Thr Gly
385 390 395

Ile Tyr Phe Ile Pro Leu Asn Asn Ala Cys Ser Gly Met Thr Ala
400 405 410

20 Val Asp Gln Glu Phe Ser Ser Leu Asp Val Tyr Asn Val Ser Leu
415 420 425

Asp Tyr Lys Leu Ser Pro Gly Ser Glu Asn Met Gly Arg Ile Asp
25 430 435 440

Ala Ile Asp Ile Ser Thr Gly Arg Thr Leu Trp Ser Ala Glu Arg
445 450 455

30 Tyr Ala Ser Asn Tyr Ala Pro Val Leu Ser Thr Gly Gly Val
460 465 470

Leu Phe Asn Gly Gly Thr Asp Arg Tyr Phe Arg Ala Leu Ser Gln
35 475 480 485

Glu Thr Gly Glu Thr Leu Trp Gln Thr Arg Leu Ala Thr Val Ala
490 495 500

Ser Gly Gln Ala Ile Ser Tyr Glu Ile Asp Gly Val Gln Tyr Val
40 505 510 515

Ala Ile Gly Arg Gly Gly Thr Ser Tyr Gly Ser Asn His Asn Arg
520 525 530

45 Ala Leu Thr Glu Arg Ile Asp Ser Thr Ala Ile Gly Ser Ala Ile
535 540 545

Tyr Val Phe Ala Leu Pro Gln Gln
550 555

50

INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 579 residues

(B) TYPE: amino acid

5 (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) ORIGINAL SOURCE:

ORGANISM: *Gluconobacter oxydans*

STRAIN: DSM 4025

10 (iv) FEATURE:

FEATURE KEY: sig peptide

POSITION: -23..-1

15 SEQUENCING METHOD: E

FEATURE KEY: mat peptide

20 POSITION: 1..556

SEQUENCING METHOD: E

25 Met Asn Pro Thr Thr Leu Leu Arg Thr Ser Ala Ala Val Leu Leu
-20 -15 -10

30 Leu Thr Ala Pro Ala Ala Phe Ala Gln Val Thr Pro Ile Thr Asp
-5 1 5

Glu Leu Leu Ala Asn Pro Pro Ala Gly Glu Trp Ile Asn Tyr Gly
10 15 20

35 Arg Asn Gln Glu Asn Tyr Arg His Ser Pro Leu Thr Gln Ile Thr
25 30 35

40 Ala Asp Asn Val Gly Gln Leu Gln Leu Val Trp Ala Arg Gly Met
40 45 50

45 Glu Ala Gly Ala Val Gln Val Thr Pro Met Ile His Asp Gly Val
55 60 65

Met Tyr Leu Ala Asn Pro Gly Asp Val Ile Gln Ala Leu Asp Ala
70 75 80

5 Gln Thr Gly Asp Leu Ile Trp Glu His Arg Arg Gln Leu Pro Ala
85 90 95

10 Val Ala Thr Leu Asn Ala Gln Gly Asp Arg Lys Arg Gly Val Ala
100 105 110

15 Leu Tyr Gly Thr Ser Leu Tyr Phe Ser Ser Trp Asp Asn His Leu
115 120 125

20 Ile Ala Leu Asp Met Glu Thr Gly Gln Val Val Phe Asp Val Glu
130 135 140

25 Arg Gly Ser Gly Glu Asp Gly Leu Thr Ser Asn Thr Thr Gly Pro
145 150 155

30 Ile Val Ala Asn Gly Val Ile Val Ala Gly Ser Thr Cys Gln Tyr
160 165 170

35 Ser Pro Tyr Gly Cys Phe Ile Ser Gly His Asp Ser Ala Thr Gly
175 180 185

40 Glu Glu Leu Trp Arg Asn His Phe Ile Pro Gln Pro Gly Glu Glu
190 195 200

45 Gly Asp Glu Thr Trp Gly Asn Asp Phe Glu Ala Arg Trp Met Thr
205 210 215

50 Gly Val Trp Gly Gln Ile Thr Tyr Asp Pro Val Thr Asn Leu Val
220 225 230

55 Phe Tyr Gly Ser Thr Gly Val Gly Pro Ala Ser Glu Thr Gln Arg
235 240 245

60 Gly Thr Pro Gly Gly Thr Leu Tyr Gly Thr Asn Thr Arg Phe Ala
250 255 260

65 Val Arg Pro Asp Thr Gly Glu Ile Val Trp Arg His Gln Thr Leu
265 270 275

70 Pro Arg Asp Asn Trp Asp Gln Glu Cys Thr Phe Glu Met Met Val
280 285 290

75 Ala Asn Val Asp Val Gln Pro Ser Ala Glu Met Glu Gly Leu Arg
295 300 305

80 Ala Ile Asn Pro Asn Ala Ala Thr Gly Glu Arg Arg Val Leu Thr
310 315 320

Gly Ala Pro Cys Lys Thr Gly Thr Met Trp Ser Phe Asp Ala Ala
325 330 335

5 Ser Gly Glu Phe Leu Trp Ala Arg Asp Thr Asn Tyr Thr Asn Met
340 345 350

10 Ile Ala Ser Ile Asp Glu Thr Gly Leu Val Thr Val Asn Glu Asp
355 360 365

Ala Val Leu Lys Glu Leu Asp Val Glu Tyr Asp Val Cys Pro Thr
370 375 380

15 Phe Leu Gly Gly Arg Asp Trp Ser Ser Ala Ala Leu Asn Pro Asp
385 390 395

Thr Gly Ile Tyr Phe Leu Pro Leu Asn Asn Ala Cys Tyr Asp Ile
400 405 410

20 Met Ala Val Asp Gln Glu Phe Ser Ala Leu Asp Val Tyr Asn Thr
415 420 425

25 Ser Ala Thr Ala Lys Leu Ala Pro Gly Phe Glu Asn Met Gly Arg
430 435 440

Ile Asp Ala Ile Asp Ile Ser Thr Gly Arg Thr Leu Trp Ser Ala
445 450 455

30 Glu Arg Pro Ala Ala Asn Tyr Ser Pro Val Leu Ser Thr Ala Gly
460 465 470

Gly Val Val Phe Asn Gly Gly Thr Asp Arg Tyr Phe Arg Ala Leu
475 480 485

35 Ser Gln Glu Thr Gly Glu Thr Leu Trp Gln Ala Arg Leu Ala Thr
490 495 500

40 Val Ala Thr Gly Gln Ala Ile Ser Tyr Glu Leu Asp Gly Val Gln
505 510 515

Tyr Ile Ala Ile Gly Ala Gly Gly Leu Thr Tyr Gly Thr Gln Leu
520 525 530

45 Asn Ala Pro Leu Ala Glu Ala Ile Asp Ser Thr Ser Val Gly Asn
535 540 545

50 Ala Ile Tyr Val Phe Ala Leu Pro Gln
550 555

See B1 cont

INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 82 bases
(B) TYPE: nucleotide
(C) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (iii) ORIGINAL SOURCE: synthetic oligonucleotide

5 CATGAAAATA AAAACAGGTG CACGCATCCT CGCATTATCC GCATTAACGA 50

10 CGATGATGTT TTCCGCCTCG GCTCTCGCCC AG 82

INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 83 bases
(B) TYPE: nucleotide
(C) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (iii) ORIGINAL SOURCE: synthetic oligonucleotide

15 GTTACCTGGG CGAGAGCCGA GGCGGAAAAC ATCATCGTCG TTAATGCGGA 50

20 TAATGCGAGG ATGCGTGCAC CTGTTTTAT TTT 83

INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 residues

(B) TYPE: amino acid

5 (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) ORIGINAL SOURCE: *E. coli*

(iv) FEATURE:

FEATURE KEY: sig peptide

10 POSITION: 1..26

FEATURE METHOD: S

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu
1 5 10 15

15 Thr Thr Met Met Phe Ser Ala Ser Ala Leu Ala Gln
20 25 27

20 INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 bases

(B) TYPE: nucleotide

(C) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA

(iii) ORIGINAL SOURCE: synthetic oligonucleotide

30 GTTAGCGCGG TGGATCCCCA TTGGAGG 27